PME 43

CONFERENCE OF THE INTERNATIONAL GROUP FOR
THE PSYCHOLOGY OF MATHEMATICS EDUCATION

07-12 July 2019 **Pretoria South Africa**

Proceedings

43rd Annual Meeting of the International Group for the Psychology of Mathematics Education

VOLUME 2

Research Reports (A-K)

Editors:
Mellony Graven, Hamsa Venkat, Anthony A Essien and Pamela Vale

Psychology in Mathematics Education
(PME) Conference
7-12 July 2019
University of Pretoria,
South Africa

Website: pme43@up.ac.za

Cite as:

Graven, M., Venkat, H., Essien, A. & Vale, P. (Eds). (2019). *Proceedings of the 43rd Conference of the International Group for the Psychology of Mathematics Education* (Vol 2). Pretoria, South Africa: PME.

Website: https://www.up.ac.za/pme43

Proceedings are also available on the IGPME website: http://www.igpme.org

ISBN (Volume 2, print): 978-0-6398215-2-8

ISBN (Volume 2, electronic): 978-0-6398215-3-5

Printed by Minute Man Press, Hatfield

Logo designed by GA Design

TEACHERS AND STANDARDIZED ASSESSMENTS IN MATHEMATICS: AN AFFECTIVE PERSPECTIVE

Pietro Di Martino, Giulia Signorini

Department of Mathematics, Università di Pisa

Standardized assessments in mathematics have an increasing relevance in the educational debate and, often, they heavily affect educational policies. Specifically, the framework and the items of standardized assessments suggest what is considered relevant as an outcome of mathematics education at a certain school level. The strength and the quality of the educational impact of standardized assessments seem to depend heavily on teachers' affective reactions to standardized assessment; however, studies focused on this issue are very rare: what are teachers' attitudes towards the standardized assessments and their effects? In this frame, we carried out a large qualitative research to investigate teachers' attitudes in the Italian context.

INTRODUCTION

National standardized tests have been officially introduced in several countries on the wave of the most famous international programmes for student assessments (PISA and TIMSS). The framework and the items of the standardized assessments suggest what is considered relevant as outcomes of mathematics education at a certain school level. Therefore, more or less in an explicit way, standardized assessments intend to impact directly not only the educational reform promoted by politicians (Breakspear, 2012), but also, at the classroom level, teachers' educational choices. The test results are often used to assess the general quality of an educational system and this has often triggered a dispiriting horse race between countries, but also between schools at a national level. Nowadays, the growing relevance of standardized assessments in mathematics in the educational debate is a fact (Kanes, Morgan & Tsatsaroni, 2014). Several studies from different traditions in educational research have focused on the reliability of standardized assessments' results, discussing what such assessments really assess, to which degree they may be viewed as didactically consistent with official curricula (Bodin, 2005), and discussing their equity (Boaler, 2003). Another line of research focuses on students' performance, analyzing, in particular, students' errors (Wijaya et al., 2014) and interpreting factors affecting students' performance (Papanastasiou, 2000). Even though it has been shown that teacher affect heavily influences instruction and learning (Jacobson & Kilpatrick, 2015), it is curious to observe that the majority of these studies has a cognitive and epistemological perspective and little regard is paid to the variable 'teachers': only recently, Di Martino and Baccaglini-Frank (2017) introduced and discussed the concept of "developmental potential" of standardized tests, seen as the educational opportunities for teachers offered by a critical approach to tests.

2019. In M. Graven, H. Venkat, A. Essien & P. Vale (Eds.). *Proceedings of the 43rd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 185-192). Pretoria, South Africa: PME.

In this frame, we carried out a large qualitative research to investigate teachers' attitudes towards the Italian test promoted by the National Institute for the Assessment of the Educational and Instructional System (INVALSI). In this paper, we focus on the following research questions: what are teachers' emotions towards the INVALSI test and what are the declared reasons to justify these feelings? Are there significant differences between teachers of different school levels?

METHOD AND RATIONALE

The context

Every year INVALSI develops and administers in May a census test for grades 2, 5, 8, 10. The number of items and the time granted vary depending on the school level. INVALSI shares the PISA framework and it designs the test items according to the official Italian National Standards. Despite this, there is a harsh debate because there is a unique test for grade 10, that is not differentiated for scientific high schools and professional institutes. The results do not affect the students' marks except for at grade 8, where each student's mark is entered (at least up to 2018) as part of the exam marking. In July, INVALSI sends back to the schools a quantitative report with the average score of each school class, and the comparison of the average scores of other schools with similar characteristics (numbers of students, social environment, etc.).

The collection of data

The choices of the research instruments and how to use them are not neutral. We developed our research within the more recent trend on affective factors in mathematics education. In particular, we refer to the shift from a normative approach – aimed to explain affective phenomena through general rules based on a cause–effect scheme – to an interpretive one – aimed to interpret the phenomena, "making sense of the world" (Di Martino & Zan, 2015). A movement towards the development and use of qualitative methods (essays, diaries, written open questionnaires, oral interviews) emerges in research on affect (and more in general in mathematics education). We developed our research within this frame.

In the first stage of our research we developed an online questionnaire and promoted the teachers' participation (on a voluntary and anonymous basis) with the help of the Italian regional education offices. In the last part of the questionnaire, participants had the possibility of sharing their e-mail address to participate to a non-anonymous semi-structured interview. This interview was developed to zoom into certain issues that emerged from the questionnaire answers. We were aware that we would not get a statistical sample in this way; however, our goal was to describe, interpret and understand a phenomenon and not to, in some sense, *measure* it. Therefore, we believe that having a convenience sample is not a limit. Participation exceeded all expectations: we collected 1964 questionnaire replies (very well distributed among the three school levels, see Table 1) and 798 participants agreed to participate in the second interview-stage. On one hand, this exceptional participation confirmed the teachers' interest in making their voices heard about this topic; on the other hand, it raised the problem of

the analysis of such a large amount of qualitative data, which needed to be very structured.

Conference Year	Number of questionnaire replies
Primary school (1-5)	635
Middle school (6-8)	643
High school (9-13)	681
Others	8

Table 1: Distribution of questionnaire replies among school levels.

The online questionnaire includes 28 questions divided into screens: background information (4), emotions (4), view on INVALSI items (5), perceived goals of the national assessment (3), strengths and weaknesses of the evaluation system (4), relationship between test and didactical practices (4), view on the evaluation system (4). Excluding the section about background information, the questions were mostly (15 out of 24) open questions; indeed, this approach allows to "catch the authenticity, richness, depth of response, honesty and candor which are the hallmarks of qualitative data" (Cohen et al., 2007, p. 249). Psychologically central beliefs and emotions of the respondent emerge, differently from what happens with the traditional scales, where the respondent has only to express a degree of agreement with respect to items chosen by others, which may be not relevant for him/her.

The analysis of data

We approached the data within the social constructionist paradigm of the grounded theory. In their original work, Glaser and Strauss (1967) introduce grounded theory as an inductive approach of research (theory and focus have to emerge from data) that is at odds with "grand" theory, where data are used to verify a pre-determined theory. The radical grounded theory conflicts with the current view that researcher's knowledge, interests, values, attitudes, emotions and beliefs not only strongly affect the research in all its development, but constitute an added value. Strauss himself understand the limits of the original and radical idea of grounded theory, stating: "the final theory that is constructed through grounded in data is a representation of both participant and researcher" (Strauss & Corbin, 1998, p. 35). This consideration led to the development of different versions of the grounded theory, including the social constructionist one:

> *Rather than assuming that theory emerges from data, constructionists assume that researchers construct categories of the data* [...] *Social constructionists disavow the idea that researchers can or will begin their studies without prior knowledge and theories about their topics. Rather than being a tabula rasa, constructionists advocate recognizing prior knowledge and theoretical preconceptions and subjecting them to rigorous scrutiny.* (Charmaz, 2008, p. 402-404)

As for the present study, the context is quite well known by researchers: the INVALSI test provoked a spirited debate since its introduction; teachers and students often stand up against them and during the test days strikes are not rare. The main purpose of the present study was to understand the phenomenon and its causes. In order to do that, one issue was finding ways to reduce the copious amounts of data into manageable and comprehensible proportions. This was done through a process of coding that constantly took shape (the codes initially introduced were often refined during the analysis) and that ended when *theoretical saturation* was reached:

> *In constant comparison the researcher compares the new data with existing data and categories, so that the categories achieve a perfect fit with the data. New and emergent categories are developed in order to be able to incorporate and accommodate data in a good fit.* (Cohen et al., 2007, p. 493)

This coding process allowed to detect frequencies (which codes were occurring most commonly) and patterns (which codes occurred together).

RESULTS AND DISCUSSION

In this paper we will discuss the analysis of the answers to Question Q5 ("*What emotion comes to your mind when you think of INVALSI?*") and Q6 ("*What are the reasons for this emotion?*"), designed to investigate teachers' emotions and the declared reasons for these emotions. The tag cloud in Figure 1 summarises the wide range of labels used and the number of their occurrences in the answers to Q5. '*Anxiety*' is the label with the largest number of occurrences (240: more than 12 per cent of the large sample), followed by '*curiosity*' (182). More in general, the quantitative analysis of the data confirms a clear prevalence of labels that describe negative emotions towards the test (the 57% of the total). There are also very interesting qualitative differences: we can recognize a category of recurring emotions (*curiosity, interest, useless, boredom*) that, in some sense, expresses the judgment about the relevance of the test; but, in the negative case, we can also recognize a category that describes a strong emotional involvement (*anxiety, stress, anger, frustration, apprehension*).

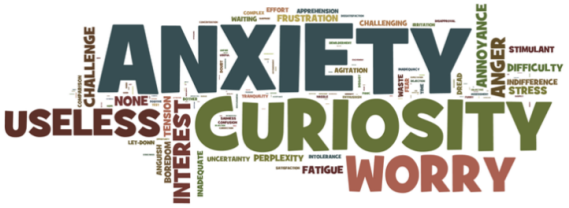

Figure 1: Tag cloud for Q5

The analysis of the answers to Q6, developed using the methodology described above, allowed to recognize three main and strictly interrelated categories of reasons for expressed emotions. These categories are related to judgements about the *adequacy, use* and *educational effects* of this kind of standardized assessment.

Adequacy

The adequacy is assessed at different levels, in terms of: mathematical content; item typology; general test structure and setting; *equity* of the test.

Mathematical content. Negative emotions are associated to the belief that INVALSI mathematics item content differs too heavily from the implemented curriculum ("*The arguments do not fit the actual curricula*", "*Items include questions about probability and statistics that are addressed only peripherally in primary school*"). The point is that INVALSI items are actually linked to the official curriculum (for example, probability and statistics are included in the Standards for primary school). The gap between Standards and implemented curriculum appears very clearly: in particular, at the primary school level, where the absence of a final examination probably gives greater freedom in the choice of the implemented curriculum.

Item typology. INVALSI items challenge students to solve stimulating non-routine problems: here again, the gap emerges between the problems proposed in the textbooks, and ultimately with the implemented curriculum. This fact is used as a reason to justify both positive and negative emotions. The positive emotions related to this fact implicitly highlight the negative opinion about the *curricular* problems ("*I like the types of problem used by INVALSI because they are different from the one used in classroom: they also seem closer to students' context of life*"). Vice versa, reasons for negative emotions highlight the belief that a *good* standardized assessment should adapt the items to the traditionally implemented curriculum ("*They need to design a test that is in line with what teachers do in classrooms everyday*") or to the textbooks' style ("*The problems in the test are completely different from those contained in the textbooks*") rather than to the Standards.

General test structure and setting. This is a recurrent issue to justify negative emotions towards the test: multiple choice tests are considered inadequate for assessing mathematics competence; primary school teachers label the children's lack of opportunity to ask the teacher for help as unnatural; the allotted time period is considered inadequate for coping with challenging items. Interestingly, we found 267 occurrences of the time factor in the analysis of Q6: 258 of them are related to a negative emotion and only 9 of them to a positive emotion! In the latter case, teachers underline that the setting is similar to other selections that children will have to face in their lives and for this reason it can be formative.

Equity of the test. A number of criticisms accuse the test and its scoring system as being unequal because they are intended to assess all students equally, without taking into account students' social backgrounds and their starting points. In particular, secondary teachers underline the difference between students from high schools and professional

institutes, while primary and middle school teachers report the difficulties, also at an affective level, for students with special needs.

Use

A widespread perception emerges about the test results being used to assess teachers' efficacy. While official documents make it explicit that the INVALSI test will not be used to evaluate teacher (as it is, instead, in other countries), it is true that, locally, teachers with the worst results are often put under pressure by their principals. In this case, it is very interesting to look at the different reactions of primary and middle school teachers on one side, and secondary teachers on the other. The former uniformly criticize the fact that they are being indirectly assessed; the latter accept the possibility of being assessed, but complain about the students' scarce motivation to perform well and about the absence of differentiations ("*I think it is not right that results obtained from different classes, different schools, different regions, are used to draw conclusions about the validity of a teacher*").

Educational effects

This category is strictly related to the previous one: teachers believe that the tests have educational effects, especially because their use goes beyond simply assessing students' mathematical competence. In some sense, there is the widespread belief that the introduction of the test has really affected teachers' practice. While some secondary school teachers appreciate that INVALSI *forces* the teachers of the lower school levels to cover all the contents in the Standards and to propose non routine mathematical problems, in general primary school teachers criticize the educational effects of the test. They feel the pressure of having to cover all the topics included in the curriculum, regardless of the specific needs of their classes ("*You should not try to standardize teaching, you should encourage teachers to take into account the specific needs of their classroom*"): they believe that this system forces to privilege quantity rather than quality in education. On the other hand, the main criticism advanced by middle and secondary school teachers is the risk of promoting a method of education focused on preparing students for a standardized test, the so called teaching to test approach ("*The tests are likely to interfere excessively with the teaching, inducing an ad hoc training*").

CONCLUSION

The quantitative analysis of the data collected in our study highlights two facts. On one hand, the exceptional voluntary participation proves the teachers' interest towards the issue of standardized assessment of students' mathematical competence. On the other hand, the supposed prevalence of negative feelings towards the national standardized assessment of students' mathematical competence is confirmed. However, what is really important is the qualitative analysis of the different reasons for the teachers' declared emotions. What emerges is a complex picture that includes positions of

principle against the standardized assessments and their uses, but also more specific criticism towards the design of the test. In this contribution, we discussed only a small part of the data collected in the whole research project, but, for example, the data collected with the closed question Q25 ("*What would you do with the INVALSI test?*") show that *only* the 27% of respondents stated the desire to suppress the standardized assessment: the 60% of the sample would maintain the test, recognizing its role, but asking for significant amendments.

At the end of our study, we are even more convinced it is fundamental to consider teachers' affect and to listen teachers' voices about standardized assessment: teachers are one of the key players in the educational context and their active involvement is needed to improve the tests and to exploit their informational and developmental potential (Di Martino & Baccaglini-Frank, 2017). As Ponte et al. argue:

> The reasons for studying the views and attitudes of teachers are grounded in the assumption that these have a significant influence on their thinking and actions. Views and attitudes act as a sort of filter. (Ponte et al., 1994, p. 347)

References

Bodin, A. (2005). *What does PISA really assess? What it doesn't? A French view*. Paper presented at the Joint Finnish-French Conference: Teaching mathematics: beyond the PISA survey. Available to: www.apmep.fr/IMG/pdf/Com_PISA_FF_English.pdf

Boaler, J. (2003). When learning no longer matters - Standardized testing and the creation of inequality. *Phi Delta Kappan*, *84*(7), 502-506.

Breakspear, S. (2012) The policy impact of PISA: an exploration of the normative effects of international benchmarking in school system performance. *OECD Education Working Papers*, 71. Paris: OECD Publishing. http://dx.doi.org/10.1787/5k9fdfqffr28-en

Charmaz, K. (2008). Constructionism and the Grounded Theory Method. In J. Holstein & J. Gubrium (Eds.), *Handbook of Constructionist Research* (pp. 397-412). New York: The Guiltord Press.

Cohen, L., Manion, L. & Morrison, R. (2007). *Research methods in education*. London: RoutledgeFalmer.

Di Martino, P. & Baccaglini-Frank, A. (2017). Beyond performance results: analyzing the informational and developmental potentials of standardized mathematics test. *For the Learning of Mathematics*, *37*(3), 39-44.

Di Martino, P. & Zan, R. (2015). The Construct of Attitude in Mathematics Education. In B. Pepin & B. Roesken-Winter (Eds.), *From beliefs to dynamic affect systems in mathematics education. Advances in Mathematics Education*. New York: Springer.

Glaser, B. & Strauss, A. (1967). *The discovery of grounded theory. Strategies for qualitative research*. Chicago: Aldine.

Jacobson, E. & Kilpatrick, J. (2015). Understanding teacher affect, knowledge, and instruction over time: an agenda for research on productive disposition for teaching mathematics. *Journal of Mathematics Teacher Education*, *18*(5), 401-406.

Kanes, C., Morgan, C. & Tsatsaroni, A. (2014). The PISA mathematics regime: knowledge structures and practices of the self. *Educational Studies in Mathematics, 87*(2), 145-165.

Papanastasiou, C. (2000). Internal and external factors affecting achievement in mathematics: Some findings from TIMSS. *Studies in Educational Evaluation*, 26, 1-7.

Ponte, J., Matos, J., Guimarães, H., Leal, L. & Canavaro, A. (1994). Teachers' and students' views and attitudes towards a new mathematics curriculum: A case study. *Educational Studies in Mathematics, 26*(4), 347-365.

Strauss, A. & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd edition). Thousand Oaks: Sage Publications

Wijaya, A., Van den Heuvel-Panhuizen, M., Doorman, M., & Robitzsch, A. (2014). Difficulties in solving context-based PISA mathematics tasks: An analysis of students' errors. *The Mathematics Enthusiast*, *11*(3), 555-584.